

Para citar este artículo: Celis Bueno, C. (2022). Automatizar el sentido común: moderación de contenidos en el capitalismo de plataformas. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, 15(2). 1-15. <https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.10930>

AUTOMATIZAR EL SENTIDO COMÚN: MODERACIÓN DE CONTENIDOS EN EL CAPITALISMO DE PLATAFORMAS

Automating Common Sense: Content Moderation in Platform Capitalism

Automatizando o senso comum: moderação de conteúdo no capitalismo de plataforma

Claudio Celis Bueno, University of Amsterdam (Países Bajos)

claudiocelisbueno@gmail.com

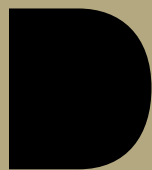
Recibido: 03 de agosto de 2021

Aceptado: 06 de septiembre de 2021

Fecha de prepublicación: 12 de marzo de 2022

RESUMEN

El presente artículo desarrolla una reflexión sobre el proceso de automatización de la moderación de contenidos en redes sociales. Para ello, el artículo se divide en dos partes. La primera se concentra en las consecuencias políticas de delegar en un algoritmo el juicio que distingue entre lo que es y no es "publicable". Se trata de comprender las consecuencias políticas generadas por el proceso de automatización de lo que cierta tradición filosófica ha llamado "el sentido común". La segunda se concentra en la dimensión económico-política detrás de este proceso de automatización. A partir de una metodología crítica (posmarxista), se examinó la relación entre trabajo y valor que subyace a esta "automatización del sentido común". Desde la perspectiva de una crítica de la economía política, se sugiere que la automatización de la moderación de contenidos depende de una apropiación y una



explotación de lo “común”. En este sentido, la moderación de contenidos y su automatización funcionan como caso ejemplar para analizar los conceptos de trabajo y de valor en lo que se ha llamado “extractivismo de datos” y “capitalismo de plataformas”.

Palabras clave: redes sociales; capitalismo de plataformas; extractivismo de datos; trabajo; algoritmos; inteligencia artificial; política; Facebook.

ABSTRACT

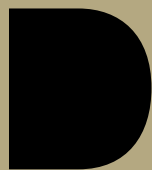
This article reflects on the process of automating content moderation on social networks. To this end, the article is divided into two parts. The first section focuses on the political consequences of delegating the judgment of whether something is “publishable” to an algorithm. It is about understanding the political consequences of the automation process of what certain philosophical tradition has called “common sense.” The second part focuses on the economic-political dimension behind this automation process. From a critical (post-Marxist) methodology, we examine the relationship between labour and value that underlies this “automation of common sense.” From the perspective of a critique of political economy, it is suggested that the automation of content moderation depends on an appropriation and exploitation of the “common”. In this sense, content moderation and automation serve as an exemplary case for analyzing the concepts of labour and value in the so-called “data extractivism” and “platform capitalism.”

Keywords: Social media; platform capitalism; data extractivism; labour; algorithms; artificial intelligence; politics; Facebook.

RESUMO

Este artigo desenvolve uma reflexão sobre o processo de automação da moderação de conteúdo em redes sociais. Para isso, o artigo está dividido em duas partes. A primeira parte enfoca as consequências políticas de delegar a um algoritmo o julgamento que distingue entre o que é e o que não é “publicável”. Trata-se de compreender as consequências políticas geradas pelo processo de automação do que uma certa tradição filosófica chamou de “bom senso”. A segunda parte enfoca a dimensão político-econômica por trás desse processo de automação. Numa metodologia crítica (pós-marxista), examina-se a relação entre trabalho e valor subjacente a esta “automação do senso comum”. Na perspectiva de uma crítica à economia política, sugere-se que a automação da moderação de conteúdos depende de uma apropriação e exploração do “comum”. Nesse sentido, a moderação e automação de conteúdo funcionam como um caso exemplar de análise dos conceitos de trabalho e valor no denominado “extrativismo de dados” e “capitalismo de plataforma”.

Palavras-chave: mídia social; capitalismo de plataforma; extrativismo de dados; trabalho; algoritmos; inteligência artificial; política; Facebook.



Moderación de contenidos en redes sociales

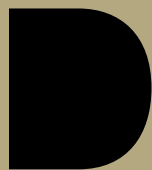
La moderación de contenidos es la compleja tarea de filtrar los contenidos publicados diariamente por los miles de millones de usuarios en redes sociales para evitar “actividades abusivas”. Qué entendamos exactamente por actividades abusivas depende de las normativas de cada plataforma, pero en términos generales se refiere a publicaciones que promueven la violencia, los discursos de odio y la discriminación, y que atentan contra la dignidad humana; también se incluyen fenómenos como las noticias falseadas (*fake news*), la usurpación de identidad, la pornografía y los contenidos que violan el derecho de autor (Koebler & Cox, 2018).

El presente artículo utilizará como ejemplo el caso de Facebook, que ha sido uno de los más polémicos y mediáticos en relación con este tema (Koebler & Cox, 2018). La moderación del contenido de los más de dos mil millones de usuarios de Facebook genera una tensión entre una plataforma que asegura que la “libertad de expresión” define su supuesto “espíritu” pero, al mismo tiempo, es capaz de bloquear contenidos “abusivos” que circulan por ella. La declaración de políticas de moderación de contenido de Facebook (2021) dice:

Cada día, las personas usan Facebook para compartir sus experiencias, conectarse con amigos y familiares, y crear comunidades. Ofrecemos nuestro servicio a más de 2000 millones de personas para que se expresen libremente en diversos países y culturas y en decenas de idiomas. [...] Reconocemos la importancia de Facebook como un espacio donde las personas se sientan libres para comunicarse y, por ello, tomamos muy en serio nuestra responsabilidad de mantener nuestro servicio libre de abusos. [...] La misión de crear comunidades y unir más al mundo depende de la posibilidad de las personas para compartir diferentes perspectivas, experiencias, ideas e información. [...] Si bien nuestro compromiso con la libertad de expresión es de suma importancia, reconocemos que internet crea cada vez más oportunidades para cometer abusos. Por este motivo, si limitamos la expresión es porque buscamos proteger uno o más de los siguientes valores: autenticidad, seguridad, privacidad, y dignidad.

La moderación de contenido es una suma de millones de microdecisiones que filtran la actividad en línea, guiadas por las “normas” establecidas por la propia empresa. En este sentido, la moderación de contenidos se diferencia de las macrodecisiones de censura en redes sociales, como aquella de suspender la cuenta del expresidente Donald Trump (tema que abre toda otra serie de problemas que no serán tratados aquí). Como se mencionó más arriba, las “normas” que guían las microdecisiones de moderación de contenidos abarcan una serie de áreas como violencia, pornografía, terrorismo, discriminación, actividades ilegales, propiedad intelectual, autenticidad y privacidad, entre otras. Si bien estas normas se encuentran en permanente revisión (Facebook, 2021), su definición e implementación han sido fuertemente criticadas en los últimos años (Patel & Hecht-Felella, 2021). Las principales críticas se concentran en cuestiones tales como algoritmos poco confiables, reglas poco claras o muy generales, falta de contexto para las decisiones, diferencias lingüísticas y poca sensibilidad a la proporcionalidad de las “faltas” (Patel & Hecht-Felella, 2021).

No obstante, el problema no son solo las normas definidas por Facebook. Nos encontramos ante un fenómeno de una escala tal que implica un problema técnico que a su vez se traduce en problemas políticos, éticos y económicos. Como proponen Koebler y Cox (2018), la moderación de contenidos es uno de los desafíos que implica mayor cantidad de “trabajo” y de “problemas logísticos complejos” que haya tenido que enfrentar Facebook, considerando

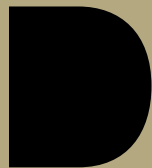


los dos mil millones de usuarios que publican diariamente en más de cien idiomas. La escala de la cantidad de información que debe ser filtrada implica que la moderación de contenido no puede ser realizada solamente por “filtros humanos”. Como ya decía Herbert Simon (1971), el problema de la “sociedad de la información” no es la información, sino la atención necesaria para procesarla (Celis Bueno, 2017). En este sentido, Koebler y Cox (2018) han mostrado que, en el 2018, para los dos mil millones de usuarios de Facebook, se requería que cada moderador humano moderara “más de 10 millones de publicaciones potencialmente censurables” a la semana. La escala del fenómeno ha empujado a las plataformas de redes sociales a implementar algoritmos de inteligencia artificial (IA) para automatizar la tarea de moderación de contenido (Gillespie, 2020). Esto, a su vez, abre toda otra serie de problemas, como la capacidad de estos algoritmos para contextualizar los mensajes, la sensibilidad a figuras retóricas como la ironía o el sarcasmo, las ambigüedades lingüísticas, etc. Si el filtro algorítmico es demasiado específico (*over-fitting*), habrá muchos casos que violan las normas de publicación y que este no será capaz de identificar. Por el contrario, si el filtro es demasiado general (*under-fitting*), se filtrarán más mensajes de los que se debiese, eliminando incluso aquellos que no entran en conflicto con las reglas de publicación de la página, lo que pondría en riesgo la supuesta “libertad de expresión” de la plataforma (como ejemplo véase Langran, 2021).

Por un lado, entonces, existe la necesidad práctica de utilizar filtros de IA para la moderación de contenidos. Por el otro, existen limitaciones técnicas y políticas de la IA para resolver satisfactoriamente esta compleja tarea. Por ello, las plataformas de redes sociales han optado por implementar un modelo “híbrido” en el cual se combinan tres elementos: denuncias de los propios usuarios de la plataforma, filtros de IA y decisiones humanas realizadas por trabajadores de la plataforma (en la mayoría de los casos subcontratados). La combinación de estos tres elementos ilustra de manera ejemplar aquello que Ekbja y Nardi (2017) han llamado “*heteromation*”: la extracción de valor a través de trabajo humano que opera de manera invisible detrás de la automatización y que es mal remunerado (por ejemplo: los moderadores de contenido) o simplemente no remunerado (por ejemplo: los usuarios que denuncian).

El trabajo de moderación de contenidos es un trabajo híbrido que combina IA y trabajo humano. Por lo general es tercerizado y realizado por empresas externas (Roberts, 2019). En los últimos años, un número importante de investigaciones periodísticas (Koebler & Cox, 2018; Newton, 2019; Crossfield, 2019), documentales (Block & Riesewieck, 2018) y trabajos académicos (Gillespie, 2018; Roberts, 2019) han mostrado que por lo general se trata de un trabajo precarizado, de grandes costos psicológicos y que, en muchos casos, responde a una división internacional del trabajo. Cabe mencionar que en 2020, Facebook perdió una demanda colectiva y tuvo que pagar una compensación de 52 millones de dólares a los trabajadores que sufrieron estrés postraumático como consecuencia de esta tarea (Newton, 2020).

El presente artículo se propone desarrollar una reflexión conceptual sobre las consecuencias del proceso de automatización de la moderación de contenidos. Para ello, el artículo se divide en dos partes. La primera está guiada por la pregunta acerca de las consecuencias políticas de delegar en un algoritmo el juicio que distingue entre lo que es y no es “publicable”. Como se propone desde el título de este artículo, la primera parte trata de comprender las consecuencias políticas generadas por el proceso de automatización de lo que cierta tradición filosófica ha llamado “el sentido común” (Kant, 2000; Arendt, 1992; Rancière, 1999). La segunda se concentra en la dimensión económico-política detrás de este proceso de automatización. A partir de una perspectiva posmarxista, se examinó la relación entre trabajo y valor que subyace a esta “automatización del sentido común”. Desde



la perspectiva de una crítica de la economía política, en esta parte también se sugiere que la automatización de la moderación de contenidos depende de aquello que Carlo Vercellone (2007) y Antonio Negri (2016) han llamado una apropiación y una explotación de “lo común”. En este sentido, la moderación de contenidos y su automatización funcionan como caso ejemplar para analizar los conceptos de trabajo y de valor en lo que ha sido llamado “extractivismo de datos” (Crawford, 2021) y “capitalismo de plataformas” (Srnicek, 2018).

Política y sentido común

En un texto reciente, Tarleton Gillespie (2020) se pregunta si desde el punto de vista de una política democrática y pluralista es recomendable delegar la moderación de contenidos en un algoritmo de IA. Para una importante tradición de la filosofía política, el sentido común y el acto de juzgar entre lo justo y lo injusto definen el fundamento de la política. Desde esta perspectiva, la automatización de la moderación de contenidos representaría un caso ejemplar en el cual lo que estaría en juego es precisamente la automatización de ese aspecto esencial para la política.

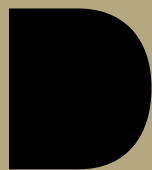
En su clásica definición, Aristóteles (1959) presenta al animal humano como un animal esencialmente político [*politikon zoon*] (p. 11). Para Aristóteles, lo que define al humano como animal político es el lenguaje, el “logos” (a veces mal traducido como “animal racional”). Aristóteles distingue el lenguaje [*logos*] de la voz [*phoné*]. Esta última, nos dice, es común a todos los animales y permite expresar el dolor y el placer. El lenguaje, en cambio, es propio de los seres humanos. El lenguaje es lo que permite distinguir entre lo justo y lo injusto. La política es el espacio donde este lenguaje es puesto al servicio del debate público entre ciudadanos, es decir, entre hombres, adultos, propietarios de esclavos, para deliberar sobre la vida justa.

En sus conferencias sobre la filosofía política de Kant de 1970, Hannah Arendt busca precisar mejor esta definición aristotélica de política. Para ello, Arendt (2003) diferencia entre el “sentido privado” y el “sentido común” (p. 120). El sentido privado es el sentido lógico, el entendimiento abstracto, que deriva conclusiones particulares a partir de reglas universales. La verdadera política, nos dice Arendt (2003), no puede derivar del pensamiento abstracto en tanto sentido privado. Precizando el argumento de Aristóteles, Arendt nos dice que el ser humano no es un animal político porque es un animal racional —porque posee el sentido privado del pensamiento lógico—, sino que el humano es un animal político porque posee el “sentido común”, es decir, la capacidad para comunicar un juicio sobre lo justo y lo injusto.

Arendt (2003) deriva su noción de “sentido común” de la lectura de la *Crítica de la facultad de juzgar* de Kant y de la diferencia que allí se establece entre juicio determinante y juicio reflexivo (p. 125). El juicio determinante pertenece al sentido privado: es la capacidad de aplicar reglas generales a casos particulares. El juicio reflexivo, en cambio, es la capacidad de extraer reglas generales de casos particulares. El sentido común es la base objetiva sobre la que se levantan los juicios reflexivos. En palabras de Kant (2000),

Por *sensus communis* ha de entenderse la idea de un sentido que es común a todos, es decir, de una facultad de juzgar que, en su reflexión, tiene en cuenta por el pensamiento (*a priori*) el modo de representación de los demás para atener su juicio, por decirlo así, a la razón total humana (§40).

Arendt nota que lo interesante de Kant es que el juicio del gusto —“esto es bello”—, en apariencia totalmente apolítico, es en efecto el fundamento de la política, ya que sienta los fundamentos de un tipo de juicio



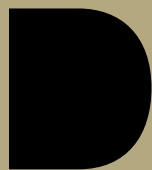
reflexivo —sentido común— no determinante —sentido privado—. Lo que define al juicio del gusto es a) su carácter reflexivo, no determinante, y b) su carácter subjetivo pero de validez universal, es decir, se trata de un juicio que se levanta sobre un sentido común que permite la comunicabilidad de juicios subjetivos. Este doble carácter del juicio de gusto —reflexividad y comunicabilidad— dependen de un sentido común. Con la relectura de Aristóteles y Kant, Arendt (2003) propone que la política, es decir, la facultad de juzgar lo justo de lo injusto, se levanta sobre las mismas dos características del juicio de gusto: la reflexividad —no es un juicio determinante que subsume lo particular a lo universal— y la comunicabilidad —se levanta sobre un sentido común compartido que posibilita la universalidad de juicios subjetivos—. El sentido común, agrega Arendt (2003) “es el sentido propiamente humano porque la comunicación, es decir, el discurso, depende de él” (p. 130). De este modo, cada vez que se juzga sobre lo justo y lo injusto, “se juzga como parte de una comunidad de humanos” (Arendt, 2003, p. 134). En síntesis, la política es la capacidad de “ponerse en el lugar del otro” en tanto miembros de una “comunidad universal”, de seres humanos, y por ello el “sentido común” y el “juicio reflexivo” constituyen las facultades humanas propiamente políticas.

Ahora bien, mientras Hannah Arendt desarrolla en detalle el argumento según el cual el “sentido común” es el fundamento de una política democrática y cosmopolita, Jacques Rancière (1999) define la política precisamente como la interrupción del sentido común. En contra del “*sensus communis*”, Rancière propondrá el “di-senso” y el “des-acuerdo”. Rancière critica la distinción entre sentido público y sentido común. Dicho en términos simples, para Rancière (1999), lo político no es lo que queda del lado del sentido común, sino el acto que decide el límite entre la esfera privada y la esfera común:

Podría decirse que, para Rancière, el *zôon politikon* que Arendt convoca ya posee títulos que lo autorizan a participar en la vida pública; que el acceso reservado al espacio público-político del que goza este actor político es en sí una señal de que solo pueden realizar su humanidad los que por su título están llamados a formar parte de la comunidad política; y que, por lo tanto, esta distribución [*partage*] de lo sensible entre lo privado y lo público es justamente lo que las luchas por la igualdad vuelven a poner incesantemente en cuestión. (Tassin, 2012, p. 45)

Tanto para Arendt (2003) como para Rancière (1999) la política es una dimensión fundamental de lo humano. Para la primera, el sentido común asegura las condiciones de igualdad en el fundamento de la política. Para el segundo, en cambio, es el “desacuerdo” sobre estas condiciones de igualdad lo que funda la verdadera política (Dikec, 2012, p. 83). Es en este sentido que Rancière (1999) distingue entre la policía —las prácticas y las instituciones que reproducen el sentido común— y la política —las prácticas que interrumpen el sentido común—. A partir de esta distinción, Rancière (1999) define la política no como la discusión sobre lo justo y lo injusto entre individuos que poseen ya la capacidad de hablar [*logos*], sino como la discusión sobre el límite entre los que pueden hablar y los que no, entre los que tienen parte y los que no tienen parte. La política no es el sentido común, que define de antemano quiénes pueden hablar, es el “disenso” y el “desacuerdo” que interrumpe el sentido común para ofrecer un nuevo reparto de las partes —entre los que tienen lenguaje [*logos*] y los que tienen voz [*phoné*]—.

¿Qué sucede cuando desplazamos estas reflexiones sobre la política hacia el fenómeno de la moderación de contenidos en redes sociales? Considerada desde el punto de vista del trabajo humano involucrado, la moderación de contenidos consiste en el desplazamiento desde la esfera pública —lo que define la política como fin en



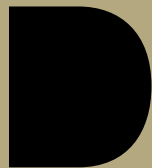
sí mismo— al espacio del trabajo privado e individual —un medio para un fin—. El juicio sobre lo justo y lo injusto es reducido a una serie de microdecisiones estandarizadas, juicios determinantes, que simplemente aplican normas generales a casos particulares. Más aún, podríamos preguntarnos si es justo o no que unos pocos decidan qué es o no es socialmente aceptado como “contenido publicable”. Bajo la misma lógica, podríamos preguntarnos si es justo o no que unos pocos, obligados a vender su fuerza de trabajo en condiciones precarias, sufran las consecuencias psicológicas de filtrar el contenido de internet para que el resto pueda disfrutar de un internet “limpio”.

Desde la perspectiva de Arendt (2003), podríamos decir que la capacidad de discernir entre lo publicable y lo no-publicable depende de un “sentido común”. Esto se pone de manifiesto en aquellos casos en que la división internacional del trabajo detrás de la moderación de contenidos se muestra insensible a las sutilezas culturales, lingüísticas y contextuales de las publicaciones que acepta o rechaza (Block & Rieseewick, 2018). Al estandarizar la decisión de lo publicable o no publicable, se estandariza el sentido común y se debilita lo que para Arendt (2003) sería la verdadera política, que es la puesta a prueba, cada vez, de ese sentido común a través de la puesta en debate de una pluralidad “inter-subjetiva” (Dikec, 2012, p. 83).

Por último, desde la perspectiva de Rancière (1999), podemos decir que la moderación de contenidos es un juicio sobre lo que puede o no puede aparecer en el ámbito de lo social. En este sentido, la moderación de contenidos cumpliría una función policial —define una determinada distribución de lo sensible—. En la medida en que asegura el “sentido común” y con ello desarticula la política, el disenso, la moderación de contenidos cumple una función policial. Dicho de otro modo, la moderación de contenidos funciona como una línea de producción que normaliza el sentido común tal como el fordismo ha normalizado el trabajo manual.

Hasta aquí la reflexión se ha referido solo a la dimensión de la moderación de contenidos en tanto “trabajo humano”. ¿Qué sucede cuando esta tarea es sometida a procesos de automatización? Una clave importante para leer este fenómeno la ofrece Hito Steyerl (2016). Este autor lee la tesis de Rancière utilizando las categorías de “ruido e información”. El ruido equivaldría a la voz [*phoné*], mientras que la información equivaldría al lenguaje [*logos*]. Para Steyerl (2016), las tecnologías de la información generan “un mar de datos” en el cual nos estamos ahogando. A diferencia de la tesis de Herbert Simon, sin embargo, Steyerl (2016) nos dice que la única manera de navegar este mar de datos y no ahogarnos en él no es ya la atención humana, sino las máquinas de reconocimiento de patrones; es decir, la automatización del proceso que distingue entre ruido e información, entre voz y lenguaje. Usando los términos de Rancière, Steyerl (2016) sostiene que los algoritmos de aprendizaje maquínico, es decir, esas sofisticadas máquinas que aprenden a identificar patrones, al distinguir entre ruido e información, automatizan el acto mismo que funda la política. Dicho de otro modo, los algoritmos de aprendizaje maquínico permiten automatizar la distinción que otrora pertenecía al dominio exclusivamente humano de la política y del lenguaje (Celis & Schultz, 2021).

Leída de este modo, la automatización de la moderación de contenidos puede ser interpretada como una automatización del juicio y del sentido común. Ahora bien, una tesis posible consiste en sostener que esta automatización es posibilitada técnicamente por un proceso previo de estandarización del sentido común. Tal como la división y estandarización del trabajo es precondition de su automatización (Marx, 2004), la estandarización del juicio y del sentido común en la moderación de contenidos es precondition de su sustitución algorítmica. Más aún, los algoritmos de aprendizaje maquínico, al ser entrenados con gigantescas bases de datos, no automatizan el sentido común en tanto facultad *a priori* y ahistórica, sino que automatizan juicios que reproducen los sesgos del



sentido común contenido en las bases de datos de entrenamiento. Más que automatizar el juicio, los algoritmos de aprendizaje automático automatizan los prejuicios contenidos en las bases de datos (para la diferencia entre juicio y prejuicio véase Arendt, 2005).

A partir de lo anterior, podemos identificar diferentes posturas. En primer lugar, Tarleton Gillespie (2020) se muestra en contra de la automatización de la moderación de contenidos, precisamente por las consecuencias políticas que conllevaría delegar en una máquina la decisión sobre lo permitido o no en el discurso público. Al mismo tiempo, sin embargo, Gillespie (2020) es consciente de que el único argumento que tendría validez para justificar la automatización de este proceso es la protección de los trabajadores humanos que realizan esta tarea —la paradoja es que esto se traduciría en la eliminación de dicho trabajo—. Un grupo importante de autoras/es ha señalado el problema del sesgo automático como un elemento estructural de los algoritmos de aprendizaje automático (Crawford, 2021; Celis & Schultz, 2020). Esto implica que utilizar esta tecnología para dichas tareas tiene siempre implícito el riesgo de que sean los prejuicios contenidos en las bases de datos los que perpetúen el sentido común a través de la decisión de qué es publicable y qué no. Eric Sadin (2019) habla de la IA como la automatización de la capacidad de juzgar e identifica en ello la muerte de la política moderna. En términos de Arendt (2003), se trataría del triunfo de la burocracia y del totalitarismo. Por último, Evgeny Morozov (2014) critica el “solucionismo tecnológico” que reemplaza la dimensión de la política por la dimensión de la técnica.

Del bando opuesto podemos encontrar a todos aquellos a favor de la automatización. Los más moderados lo justifican como mecanismo de protección del aspecto traumático del trabajo de moderación de contenidos. Encontramos de este lado también a autores que creen que el problema del sesgo y del prejuicio algorítmico se trata de un problema técnico y que para resolverlo basta mejorar los procesos de entrenamiento de los algoritmos de aprendizaje automático con bases de datos más grandes y precisas (Zarsky, 2011; 2016). Por último, cabe mencionar una serie de autores tecnoentusiastas que creen que sería precisamente en el “sentido común” donde residiría la clave para avanzar hacia una “IA general” (George et al., 2020; Pedinotti et al., 2021).

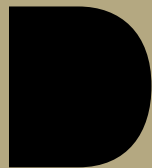
Capitalismo de plataformas y extractivismo de datos

Siguiendo la lección metodológica de Marx (1976), procederemos a abandonar

esa ruidosa esfera instalada en la superficie y accesible a todos los ojos, para dirigirnos [...] hacia la oculta morada de la producción, en cuyo dintel se lee ‘*No admittance except on business*’. Veremos aquí no solo cómo el capital produce, sino también cómo se produce el capital. Se hará luz, finalmente, sobre el misterio que envuelve la producción del plusvalor. (pp. 279-80)

Dicho de otro modo, examinaremos la automatización de la moderación de contenidos desde la perspectiva de una crítica de la economía política para identificar el lugar de esta automatización del sentido común en el proceso de valorización capitalista.

Un término que ha adquirido popularidad en los últimos años es el de “capitalismo de plataformas” forjado por Nick Srnicek (2018). Según este autor, dada la “prolongada caída de la rentabilidad de la manufactura, el capitalismo se volcó hacia los datos como un modo de mantener el crecimiento económico y la vitalidad de cara al inerte sector de la producción” (Srnicek, 2018, p. 13). Las grandes plataformas de internet (Google, Facebook,



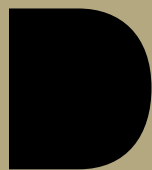
Amazon, Uber, AirBnB, etc.) deben por lo tanto ser comprendidas como parte “de una historia económica más amplia”, como “medios para generar rentabilidad” en este contexto de la caída de la tasa de ganancia del capitalismo industrial (Srniczek, 2018, p. 13). Por esta razón, Srniczek (2018) sostiene que estas plataformas estarían inaugurando un “nuevo régimen de acumulación” (p. 14).

Uno de los rasgos característicos de este nuevo régimen de acumulación sería el llamado “extractivismo de datos”, término que establece una analogía entre el manejo de información y la industria minera, que define de este modo a los datos como una materia prima que puede ser extraída, comercializada, refinada, procesada y transformada en otros bienes con valor agregado. Las plataformas constituirían así una “vasta infraestructura para detectar, grabar y analizar” datos (Srniczek, 2018, p. 42). Los datos, por su parte, aparecen como una “materia prima que debe ser extraída, y las actividades de los usuarios, la fuente natural de esta materia prima” (Srniczek, 2018, p. 42). Una importante contribución a este objetivo la han realizado Verónica Gago y Sandro Mezzadra (2017), a través de su discusión sobre la necesidad de un concepto ampliado de extractivismo para pensar el capitalismo contemporáneo.

La automatización de la moderación de contenidos representa un caso específico de extractivismo de datos. Esto se debe a que los datos son necesarios para entrenar los algoritmos de aprendizaje maquínico que hacen posible la automatización de esta tarea. El aprendizaje maquínico refiere a un tipo particular de programación en el cual un algoritmo no es diseñado paso a paso por un programador humano (como los algoritmos basados en reglas), sino que es generado de manera autónoma o semiautónoma a través de un proceso de entrenamiento. Esto se consigue a través de una función de optimización que es alimentada con una cantidad enorme de datos para extraer de ellos determinados patrones. Mientras mayor sea la cantidad de datos, mayor la efectividad del algoritmo y la complejidad de las tareas a ejecutar. Es aquí donde entra en escena el “extractivismo de datos”: el registro de información por parte de las plataformas de captura de datos ha permitido generar en el tiempo las gigantescas bases de datos que han sido luego utilizadas para el entrenamiento de algoritmos de aprendizaje maquínico. Dicho de otro modo, la acumulación de datos ha sido una de las condiciones necesarias para el desarrollo del aprendizaje maquínico.

Como decíamos, la automatización de la moderación de contenidos es un caso particular de extractivismo de datos, ya que requiere de una cantidad gigantesca de datos para entrenar a los algoritmos de aprendizaje automático. En este sentido, tanto la información que los usuarios de redes sociales que denuncian contenidos abusivos, como las decisiones de los trabajadores humanos detrás de la moderación de contenido no solo contribuyen al funcionamiento de los algoritmos en los casos más complejos, sino que además contribuyen a mejorar los procesos de entrenamiento del propio sistema de IA. Como se mencionó más arriba, se puede explicar este proceso con el término “*heteromation*” (Ekbja & Nardi, 2017): la producción de valor a través de trabajo humano invisibilizado en redes sociales que contribuye a procesos de automatización que luego reemplazarán ese mismo trabajo invisibilizado (véase también Crawford, 2021). Más aún, tal como sugiere Jan Paidos (2017), dadas las características propias del trabajo de moderación de contenidos, se trata de un caso particular de extractivismo tanto afectivo como cognitivo.

Para comprender las tesis de Srniczek (2018) y Paidos (2017), sin embargo, es necesario referirse a conceptos más generales sobre el capitalismo posindustrial desarrollados en particular por la interpretación “posoperaista” de Marx. Para los autores pertenecientes a esta línea de pensamiento posmarxista, el proceso de valorización del



capital en el capitalismo posindustrial no puede ser reducido al espacio interior de la fábrica, sino que debe ser comprendido a través del concepto de “fábrica social” (Negri, 2008). Esto implica que ya no es posible distinguir entre tiempo de trabajo y tiempo de ocio, lo cual significa, a su vez, que el tiempo abstracto de trabajo ya no puede operar como medida del valor ni, por ende, como medida de la tasa de explotación del trabajo (plusvalía). Dicho de otro modo, el capitalismo posindustrial estaría generando una crisis de la teoría del valor-trabajo de Marx (Negri 2008) y exigiendo con ello un nuevo aparato conceptual para explicar la explotación capitalista.

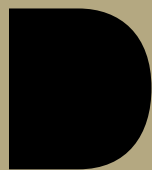
Una de las tesis centrales dentro del posoperaismo consiste en proponer que el capitalismo contemporáneo produce valor a través de la “apropiación de un común” (Negri, 2016). Si bien este común abarca también recursos naturales, para estos autores se refiere principalmente a los productos culturales que circulan en la fábrica social y que hasta ahora habrían permanecido fuera del proceso de valorización del capital (Vercellone, 2007; Negri, 2016; Hardt & Negri, 2009). Esta tesis implica dos cosas. En primer lugar, en el contexto de la fábrica social, el capital ya no organizaría el trabajo productivo, como lo hacía al interior de la fábrica industrial. En este contexto, por lo tanto, el capital ya no aparece como productivo, como fuente de un desarrollo técnico de las fuerzas productivas, sino que se limita a la apropiación y comercialización del común en espacios ya no organizados o dirigidos por este. En segundo lugar, y como consecuencia de lo anterior, Carlo Vercellone (2007; 2014) ha propuesto la tesis del “devenir renta del capital”. El capitalismo ya no sería productivo produciendo plusvalía al interior de la fábrica, sino principalmente rentista, apropiando lo común y comercializándolo sin producir nada en el proceso.

En un texto reciente, Andrea Fumagalli et al. (2018) utilizaron la tesis de la apropiación del común para explicar la producción de valor por parte de las plataformas digitales, en particular el caso de Facebook. Para estos autores, en el caso de esta red social:

el proceso descrito como devenir renta de la ganancia se vuelve evidente: Facebook no obtiene ganancia meramente de organizar el trabajo pago de sus relativamente pocos empleados, sino que extrae una renta sobre lo común producido por el trabajo no pago de sus usuarios. (Fumagalli et al., 2018, p. 35)

Desde esta perspectiva, el extractivismo de datos aparece como un importante ejemplo del devenir renta del capital y de las nuevas formas de apropiación de lo común. A través de la captura de datos, las plataformas digitales registran la actividad de los usuarios, la cual es luego procesada y comercializada. El extractivismo de datos sería un mecanismo de captura y apropiación de un común, que luego es utilizado para generar renta, ya no ganancia. Este común es registrado y luego utilizado para el entrenamiento de algoritmos de aprendizaje automático, lo que a su vez permite automatizar la producción de mercancías y acelerar el proceso de valorización del capital. El capital estaría generando una renta a partir de la expropiación de lo común a través de la extracción de datos. Más aún, este proceso estaría contribuyendo a la expulsión del trabajo humano de la esfera productiva generado por la propia automatización.

La automatización de la moderación de contenidos ilustra de manera ejemplar este vínculo entre extractivismo de datos y apropiación de lo común. Aquello que para Arendt (2003) definía el fundamento de la política (el *sensus communis*), el marco común de los ciudadanos cosmopolitas es, en este caso, apropiado a través de un proceso de extracción de datos, que luego es utilizado para automatizar las decisiones de moderación de contenidos; es decir, para distinguir entre el ruido y la información en las redes sociales. Con ello, Facebook se apropia de un “común” que luego utiliza tanto para generar valor (extraer renta), como para expulsar al trabajo humano de la esfera



productiva. En el caso de la moderación de contenidos, se trata de un híbrido entre la retroalimentación de los usuarios que denuncian alguna publicación, los filtros de IA y el trabajo de los moderadores de contenido. Los usuarios que denuncian y los moderadores de contenido no solo resuelven situaciones complejas que la IA no puede, sino que con cada microdecisión amplían las bases de datos de entrenamiento de los mismos algoritmos contribuyendo a un mejoramiento de su eficacia y acelerando la automatización del “sentido común”. Ahora bien, esta automatización del sentido común es la contracara de una “expropiación de lo común” por parte de las plataformas.

En gran medida, la tesis del “extractivismo de datos” como “apropiación de lo común” se enmarca dentro de la ya clásica formulación de Maurizio Lazzarato (1996) acerca del “trabajo inmaterial”. Para Lazzarato (1996), otro autor perteneciente al marxismo posoperaista, el trabajo inmaterial se refiere a la “actividad humana que genera la información necesaria para la producción de una mercancía” — no debe ser confundido por ende con la categoría de trabajo abstracto—. En el capitalismo contemporáneo, esto implica que muchas actividades humanas no consideradas tradicionalmente como trabajo comienzan a ser tratadas como fuente de valor. Incluso las actividades en línea de los usuarios pueden ser estudiadas como una forma de trabajo no remunerado (Terranova, 2000; Celis Bueno, 2017). Ahora bien, un modo de comprender la tesis de Lazzarato (1996) es leer la categoría de información a través del prisma de Gilbert Simondon (2016).

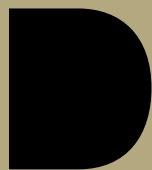
A diferencia de la teoría de la comunicación (Roman Jakobson) y de la teoría matemática de la información (Claude Shannon), Simondon (2016) define la información no como una característica interna a un mensaje, sino a partir de su capacidad para activar procesos de transformación energética. Simondon (2016) escribe:

Ser o no ser información no depende solamente de los caracteres internos de una estructura; la información no es una cosa, sino la operación de una cosa que llega a un sistema y que produce allí una transformación. La información no puede definirse más allá de ese acto de incidencia transformadora y de la operación de recepción. No es el emisor el que hace que una estructura sea información. (p. 139)

La función de información, por lo tanto, es la modificación de una realidad local —un cambio de estado energético en un sistema metaestable— a partir de una señal incidente (Simondon, 2016, p. 140). Asimismo, un receptor de información es virtualmente “toda realidad que no posee enteramente por sí misma la determinación del curso de su devenir” (Simondon, 2016, p. 140). Dicho de otro modo, todo sistema metaestable es un potencial receptor de información, es decir, un receptor de una señal incidente que gatilla una transformación de estado. Por su parte, toda señal incidente que gatille una transformación energética será considerada información.

La categoría de trabajo inmaterial de Lazzarato (1996), complementada con la categoría de información de Simondon (2016), permite: a) incluir la información como elemento central del ciclo del capital; b) identificar los procesos de acumulación de dicha información, y c) historizar el vínculo entre información y trabajo vivo en las distintas fases del capitalismo. Para finalizar este texto, revisaremos brevemente este último punto, estableciendo tres momentos de la relación entre información y trabajo vivo:

- a. En el trabajo artesanal (subsunción formal), la información necesaria para producir una mercancía se presenta como parte esencial e indisoluble del trabajo vivo. Más aún, cada vez que el artesano produce una nueva mercancía se produce un pequeño proceso de aprendizaje en el cual la información contenida en el trabajo vivo se amplifica. Esta amplificación de la información involucrada en el proceso de trabajo permanece como parte del trabajo vivo y solo puede ser expropiada por el capital en la forma de plusvalía absoluta.

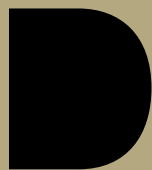


- b. En el trabajo industrial (subsunción real), en cambio, la información que antes pertenecía al artesano se objetiva en la maquinaria, que reduce el trabajo vivo a una repetición manual despojada de su dimensión informativa. En el modo de producción industrial, el capitalismo ha apropiado la información que antes pertenecía al trabajo vivo, acelerando así el proceso de producción y generando plusvalía relativa. El problema del capitalismo industrial, sin embargo, es su dificultad para modificar el proceso productivo a la luz de nuevas informaciones. Es decir, la información ingresa en la producción de mercancías, pero no es amplificada en cada nuevo ciclo ni permite ser reabsorbida de manera expedita por la línea de producción.
- c. El trabajo postindustrial surge como respuesta a esta rigidez del trabajo industrial. Su objetivo es flexibilizar el proceso de producción. Para ello, el capitalismo postindustrial (cognitivo) vuelve a incorporar en el trabajo vivo (capital variable) la dimensión informacional que antes estaba objetivada en el capital fijo. Esto permite la modificación permanente del proceso de producción, ajustándose de este modo a los flujos de un capital altamente volátil. No obstante, a diferencia del trabajo artesanal, el proceso de amplificación de la información en cada ciclo de producción no queda restringido al trabajador humano, sino que es capturado por el capital a través de las nuevas tecnologías de la información. El trabajo vivo introduce información y flexibilidad al proceso productivo las cuales son luego capturadas por las tecnologías de la información —véase, por ejemplo, el concepto de “información valorizante” de Romano Alquati (1962) y su análisis por parte de Matteo Pasquinelli (2015)—.

¿Qué sucede con la relación entre información y trabajo vivo en el caso de la automatización de la moderación de contenidos en redes sociales? Desde la perspectiva del trabajo inmaterial se podría sugerir que esta automatización depende de una apropiación del “sentido común” que acelera la separación entre información y trabajo vivo, propia del capitalismo industrial, pero que mantiene la flexibilidad del capitalismo postindustrial. Los algoritmos de aprendizaje maquínico permiten combinar una automatización radical de la producción (capital fijo) con una flexibilidad que le permite adaptarse a la velocidad y volatilidad del capitalismo contemporáneo. La relación entre extractivismo de datos y aprendizaje maquínico detrás de la automatización de la moderación de contenidos da cuenta de un proceso de captura en la forma de capital fijo de ese común que permanecía como un atributo intrínseco del trabajo vivo (“sentido común”). En este sentido, la automatización de la moderación de contenidos ilustra ejemplarmente diversas elaboraciones conceptuales tales como el “extractivismo afectivo” (Paidós, 2017); “la plataforma como fábrica” y el “trabajo fantasma detrás de la IA” (Altenried, 2020); el ya mencionado concepto de “heteromation” (Ekbia & Nardi, 2017); la relación entre datos y trabajo (Crawford, 2021), y la relación entre aprendizaje maquínico y la fábrica social (Dyer-Witheford et al., 2019).

La automatización de la moderación de contenidos en redes sociales refleja aquello que Marx (1973) definió como la paradoja propia del modo de producción capitalista: por un lado, define el trabajo humano como única fuente de valor mientras, al mismo tiempo, pone la tecnología al servicio de la expulsión de ese trabajo humano de la esfera productiva (p. 706). Actualizando esta tesis, Guattari y Negri (1990) escribieron:

Olvidemos al capitalismo y al socialismo: en su lugar encontramos una gigantesca máquina de explotación planetaria. Cada aspecto de la vida humana —trabajo, infancia, amor, vida, pensamiento, fantasía, arte— pierde toda dignidad al interior de esta gran fábrica social. Cada uno siente solo la amenaza de su



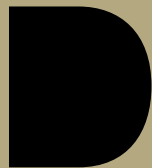
exclusión [...]. El trabajo mismo deja de ser esa promesa capaz de desarrollar el vínculo entre la humanidad y la naturaleza: hoy todos trabajan intensamente, para poder sobrevivir, y aun así no hacen más que acelerar su expulsión de esa esfera repetitiva y mecánica en la que ha devenido el trabajo.¹ (pp. 7-8)

¿No se trata de una ilustración casi literal de la situación del trabajo digital en la era contemporánea, en particular de la moderación de contenidos en redes sociales? ¿Un trabajo precarizado y de altos costos psicológicos, “repetitivo y mecánico”, que en su propia ejecución acelera la “expulsión” de todos esos restos de trabajo humano que aún habitan la “gigantesca máquina de explotación planetaria”?

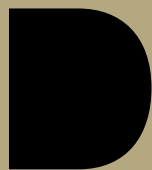
Referencias

1. Alquati, R. (1962). Composizione organica del capitale e forza-lavoro alla Olivetti. *Quaderni Rossi*, 2, 63-98.
2. Altenried, M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, 44(2), 145-158. <https://doi.org/10.1177%2F0309816819899410>
3. Arendt, H. (1992). *Lectures on Kant's political philosophy*. The University of Chicago Press.
4. Arendt, H. (2003). *Responsibility and judgment*. Schocken Books.
5. Arendt, H. (2005). *The promise of politics*. Schocken Books.
6. Aristotle. (1959). *Politics*. Harvard University Press.
7. Block, H., & Riesebeck, M. (2018) *The Cleaners*. Gebrueder Beetz Filmproduktion. [Documental, 88 mins.]. <https://gebrueder-beetz.de/en/productions/the-cleaners>
8. Celis Bueno, C. (2017). *The attention economy: Labour, time, and power in cognitive capitalism*. Rowman & Littlefield.
9. Celis Bueno, C., & Schultz, M. J. (2020). Memo Akten's learning to see: From machine vision to the machinic unconscious. *AI & Society*, 35(4), 1-11. <https://link.springer.com/article/10.1007/s00146-020-01071-2>
10. Celis Bueno, C., & Schultz, M. J. (2021). Notes on an algorithmic faculty of the imagination. *Anthropocenes: Human, Inhuman, Posthuman* (En publicación).
11. Crawford, K. (2021). *Atlas of AI*. Yale University Press.
12. Crossfield, J. (2019). *The hidden consequences of moderating social media's dark side*. <https://contentmarketinginstitute.com/cco-digital/july-2019/social-media-moderators-stress/>
13. Dikec, M. (2012). Beginners and equals: Political subjectivity in Arendt and Rancière. *Transactions of the Institute of British Geographers*, 38(1), 78-90. <https://doi.org/10.1111/j.1475-5661.2012.00508.x>
14. Dyer-Witheford, N., Mikkola Kjosien, A., & Steinhoff, J. (2019). *Inhuman power*. Pluto Press.
15. Ekbja, H., & Nardi, B. (2017). *Heteromation, and other stories of computing and capitalism*. The MIT Press.
16. Facebook. (2021). *Community Standards*. <https://www.facebook.com/communitystandards/>
17. Fumagalli, A., Lucarelli, S., Musolino, E., & Rocchi, G. (2018). El trabajo (labour) digital en la economía de plataforma: el caso de Facebook. *Hipertextos*, 6(9), 12-40. <http://revistahipertextos.org/wp-content/uploads/2015/12/1.-Fumagalli-et-al.pdf>

1 Traducción propia.



18. Gago, V., & Mezzadra, S. (2017). A critique of the extractive operations of capital: Toward an expanded concept of extractivism. *Rethinking Marxism*, 29(4), 574-591. <https://doi.org/10.1080/08935696.2017.1417087>
19. George, D., Lázaro-Gredilla, M., & Guntupalli, J. S. (2020). From CAPTCHA to commonsense: How brain can teach us about artificial intelligence. *Frontiers in Computational Neuroscience*, 14, 1-14. <https://doi.org/10.3389/fncom.2020.554097>
20. Gillespie, T. (2018). *Custodians of the Internet*. Yale University Press.
21. Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, 1-5. <https://doi.org/10.1177%2F2053951720943234>
22. Guattari, F., & Negri, A. (1990). *Communists like us*. Semiotext(e).
23. Hardt, M., & Negri, A. (2009). *Commonwealth*. Akal.
24. Kant, I. (2000). *Critique of the power of judgment*. Cambridge University Press.
25. Koebler, J., & Cox, J. (2018). The impossible job: inside Facebook's struggle to moderate two billion people. *Motherboard. Tech by Vice*. <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works>
26. Langran, C. (2021). Why was my tweet about football labelled abusive? *BBC News*. <https://www.bbc.com/news/technology-57836409>
27. Lazzarato, M. (1996). Immaterial labour. En P. Virno & M. Hardt (Eds.), *Radical thought in Italy: A potential politics* (pp. 133-147). University of Minnesota Press.
28. Marx, K. (1973). *Grundrisse: Foundations of the critique of political economy* (M. Nicolaus, Trad.). Penguin Random House.
29. Marx, K. (1976). *Capital: A critique of political economy* (Vol. 1). Harmondsworth – Penguin – New Left Review.
30. Marx, K. (2004). *Miseria de la filosofía*. Edaf.
31. Morozov, E. (2014). *To save everything, click here*. Penguin Books.
32. Negri, A. (2008). *Reflections on empire*. Polity Press.
33. Negri, A. (2016). Il comune come modo di produzione. *Euronmade*. <http://www.euronmade.info/?p=7331>
34. Newton, C. (2019). The trauma floor. *The Verge*. <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>
35. Newton, C. (2020). Facebook will pay \$52 million in settlement with moderators who developed PTSD on the job. *The Verge*. <https://www.theverge.com/2020/5/12/21255870/facebook-content-moderator-settlement-scola-ptsd-mental-health>
36. Paidos, J. (2017). Mining the mind: emotional extraction, productivity, and predictability in the twenty-first century. *Cultural Studies*, 31(2-3) 205-231. <https://doi.org/10.1080/09502386.2017.1303426>
37. Pasquinelli, M. (2015). Italian operaismo and the information machine. *Theory, Culture & Society*, 32(3), 49-68. <https://doi.org/10.1177%2F0263276413514117>
38. Patel, F., & Hecht-Felella, L. (2021). Oversight board's first rulings show Facebook's rules are a mess. *JustSecurity*. <https://www.justsecurity.org/74833/oversight-boards-first-rulings-show-facebooks-rules-are-a-mess/>
39. Pedinotti, P., Rambelli, G., Chersoni, E., Santus, E., Lenci, A., & Blache, P. (2021). *Did the cat drink the coffee? Challenging transformers with generalized event knowledge*. Ponencia presentada en la décima Joint Conference on Lexical and Computational Semantics. <https://arxiv.org/abs/2107.10922>
40. Rancière, J. (1999). *Dis-agreemet*. University of Minnesota Press.



41. Roberts, S. (2019). *Behind the screen. Content moderation in the shadows of social media*. Yale University Press.
42. Sadin, É. (2019). *Critica della ragione artificiale*. Luiss University Press.
43. Simon, H. A. (1971). Designing organizations for an information-rich world. En M. Greenberg (Ed.), *Computers, Communication, and the Public Interest* (pp. 37-72). Johns Hopkins University Press.
44. Simondon, G. (2016). *Comunicación e información*. Cactus.
45. Srnicek, N. (2018). *Capitalismo de plataformas*. Caja Negra.
46. Steyerl, H. (2016). A sea of data: Apophenia and pattern (mis)recognition. *E-Flux Journal*, 7(2), 1-14. <https://www.e-flux.com/journal/72/60480/a-sea-of-data-apophenia-and-pattern-mis-recognition/>
47. Tassin, E. (2012). De la subjetivación política. Althusser / Rancière / Foucault / Arendt / Deleuze. *Revista de Estudios Sociales*, (43), 36-49. <https://journals.openedition.org/revestudsoc/7096>
48. Terranova, T. (2000). Free labor: Producing culture for the digital economy. *Social Text*, 18(2), 35-58. <https://web.mit.edu/schock/www/docs/18.2terranova.pdf>
49. Vercellone, C. (2007). From formal subsumption to general intellect: Elements for a Marxist reading of the thesis of cognitive capitalism. *Historical Materialism*, 15(1), 13-36. <https://halshs.archives-ouvertes.fr/halshs-00263661/document>
50. Vercellone, C. (2014). Crisis de la ley del valor y devenir renta de la ganancia. Apuntes sobre la crisis sistémica del capitalismo cognitivo. En A. E. A. Fumagalli (Ed.), *La gran crisis de la economía global* (pp. 63-98). Traficantes de sueños.
51. Zarsky, T. (2011). Governmental data mining and its alternatives. *Penn State Law Review*, 116(2), 285-330. <http://www.pennstatelawreview.org/116/2/116%20Penn%20St.%20L.%20Rev.%20285.pdf>
52. Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1), 118-132. <https://doi.org/10.1177%2F0162243915605575>